PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Application Number:	10/734,486
Filing Date:	12/12/2003
Applicant(s):	William V. Da Palma et al.
Entitled:	METHOD AND PROCESS TO GENERATE REAL TIME INPUT/OUTPUT IN A VOICE XML RUN-TIME SIMULATION ENVIRONMENT
Examiner:	Joel Stoffregen
Group Art Unit:	2626
Attorney Docket No.:	BOC920030095US1 (1082-003U)

Mail Stop Amendment
Commissioner For Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

In response to the Non-Final Official Action mailed May 17, 2007 (the "Non-Final Office Action"), please enter the following amendments in the above-referenced patent application (the "Patent Application"), and further please consider the following remarks establishing the patentability of the amended claims of the Patent Application.

A listing of the Claims begins on page 2 of this paper.

Remarks/Arguments begin on page 8 of this paper.

LISTING OF THE CLAIMS

1. (Original) A method for simulating a run-time user interaction with a voice application, said method comprising the steps of:

loading a user simulation script programmed to specify simulated voice interactions with the voice application;

deriving from the voice application a nominal output, the nominal output including a text stream;

processing the user simulation script to generate a simulated output for the voice application corresponding to the nominal output;

calculating an execution time for the simulated output, the execution time being equal to a length of the text stream divided by an empirical speaking rate of a user; and

executing the simulated output at its calculated execution time in conjunction with the voice application.

2. (Original) The method of claim 1, further comprising the steps of:

processing the user simulation script to generate a simulated input for the voice application, the simulated input including a text equivalent of a pre-determined user input;

calculating an execution time for the simulated input, said execution time being equal to a length of the text equivalent of the pre-determined user input divided by the empirical speaking rate of a user; and

executing the simulated input at its calculated execution time in conjunction with the voice application.

3. (Original) A method for simulating a run-time user interaction with a voice application, said method comprising the steps of:

loading a user simulation script programmed to specify simulated voice interactions with the voice application;

deriving from the voice application a nominal output, the nominal output including an audio stream;

processing the user simulation script to generate a simulated output for the voice application corresponding to the nominal output; calculating an execution time for the simulated output based on a sampling rate and a number of samples associated with the audio stream; and

executing the simulated output at its calculated execution time in conjunction with the voice application.

4. (Original) The method of claim 3, further comprising the steps of:

processing the user simulation script to generate a simulated input for the voice application, the simulated input including an audio equivalent of a pre-determined user input to the voice application;

calculating an execution time for the simulated input based on a sampling rate and a number of samples associated with the audio equivalent; and

executing the simulated input at its calculated execution time in conjunction with the voice application.

5. (Original) A machine readable storage having stored thereon a computer program for simulating a run-time user interaction with a voice application, said computer program

comprising a routine set of instructions which when executed by a machine cause the machine to perform the steps of:

loading a user simulation script programmed to specify simulated voice interactions with the voice application;

deriving from the voice application a nominal output, the nominal output including a text stream;

processing the user simulation script to generate a simulated output for the voice application corresponding to the nominal output;

calculating an execution time for the simulated output, the execution time being equal to a length of the text stream divided by an empirical speaking rate of a user; and

executing the simulated output at its calculated execution time in conjunction with the voice application.

6. (Original) The machine readable storage of claim 5, further causing said machine to perform the steps of:

processing the user simulation script to generate a simulated input for the voice application, the simulated input including a text equivalent of a pre-determined user input;

calculating an execution time for the simulated input, said execution time being equal to a length of the text equivalent of the pre-determined user input divided by the empirical speaking rate of a user; and

executing the simulated input at its calculated execution time in conjunction with the voice application.

7. (Original) A machine readable storage having stored thereon a computer program for simulating an execution time user interaction with a voice application, said computer program comprising a routine set of instructions for causing the machine to perform the steps of:

loading a user simulation script programmed to specify simulated voice interactions with the voice application;

deriving from the voice application a nominal output, the nominal output including an audio stream;

processing the user simulation script to generate a simulated output for the voice application corresponding to the nominal output;

calculating an execution time for the simulated output based on a sampling rate and a number of samples associated with the audio stream; and

executing the simulated output at its calculated execution time in conjunction with the voice application.

8. (Original) The machine readable storage of claim 7, further causing said machine to perform the steps of:

processing the user simulation script to generate a simulated input for the voice application, the simulated input including an audio equivalent of a pre-determined user input to the voice application;

calculating an execution time for the simulated input based on a sampling rate and a number of samples associated with the audio equivalent; and

executing the simulated input at its calculated execution time in conjunction with the voice application.

9. (Original) A simulation tool for simulating a run-time user interaction with a voice application running on an application server, said tool being configured to load a user simulation script programmed to specify simulated voice interactions with the voice application, and to:

(i) first process the voice application to derive a nominal output of the voice application, the nominal output including a text stream;

(ii) second process the user simulation script to generate a simulated output for the voice application corresponding to the nominal output;

(iii) calculate an execution time for the simulated output, said execution time being equal to a length of the text stream divided by an empirical speaking rate of a user; and

(iv) execute the simulated output at its calculated execution time in conjunction with the voice application.

10. (Original) The simulation tool of claim 9, wherein the tool is further configured to:

(i) process the user simulation script to generate a simulated input for the voice application, the simulated input including a text equivalent of a pre-determined user input;

(ii) calculate an execution time for the simulated input, said execution time being equal to a length of the text equivalent of the pre-determined user input divided by the empirical speaking rate of a user; and

(iii) execute the simulated input at its calculated execution time in conjunction with the voice application.

11. (Original) A simulation tool for simulating a run-time user interaction with a voice application running on an application server, said tool being configured to load a user simulation script programmed to specify simulated voice interactions with the voice application, and to:

(i) first process the voice application to derive a nominal output of the voice application, the nominal output including an audio stream;

(ii) second process the user simulation script to generate a simulated output for the voice application corresponding to the nominal output;

(iii) calculate an execution time for the simulated output based on a sampling rate and a number of samples associated with the audio stream; and

(iv) execute the simulated output at its calculated execution time in conjunction with the voice application.

12. (Original) The simulation tool of claim 11, wherein the tool is further configured to:

(i) process the user simulation script to generate a simulated input for the voice application, the simulated input including a audio equivalent of a pre-determined user input;

(ii) calculate an execution time for the simulated input based on a sampling rate and a number of samples associated with the audio equivalent; and

(iii) execute the simulated input at its calculated execution time in conjunction with the voice application.

REMARKS

I. Overview

These remarks are set forth in response to the Non-Final Office Action. As this amendment has been timely filed within the three-month statutory period, neither an extension of time nor a fee is required. Presently, claims 1 through 12 are pending in the Patent Application. Claims 1, 3, 5, 7, 9 and 11 are independent in nature. In the Non-Final Office Action, claims 1-12 have been rejected on cited art. Specifically, claims 1 through 12 have been rejected under 35 U.S.C. § 103(a) as being unpatentable over U.S. Patent No. 6,321,198 to Hank et al. (Hank) in view of U.S. Patent No. 4,093,831 to Sharp et al. (Sharp). In response, Applicants respectfully traverse the Examiner's rejections on the art.

II. The Applicants' Invention

The Applicants have invented a system, method and apparatus for simulating a run-time user interaction with a voice application. In the Applicants' invention, a simulation environment is provided to simulate the real-time interaction between a user and a voice application. Specifically, a simulation script can be coupled to a voice application such that the script provides both simulated inputs to the application and simulated outputs from the application. The simulation script further can be configured calculate the actual execution times for the inputs and outputs to and from the voice application, which are calculated based on actual speaking rates for users. In this way, a testing environment is provided that allows the simulation of user interaction as well as the simulation of the speech technology platform, such that a developer of voice applications is no longer be dependent on human testers and speech technology and hardware to test their applications.

III. Rejections Under 35 U.S.C. § 103(a)

A. Characterization of the Cited Art

1. Hank

Hank teaches a dialogue design system for creating dialogue applications of the type understood and used by automatic speech recognition systems. The dialogue design system of Hank can be loaded into a computer and can include a Dialogue Assistant (DA). The DA includes a navigator screen for defining and connecting compartments of the dialogue application. Each compartment of a dialogue application is provided with an associated design and editing screen which enables the designer to create, build and edit compartments with associated sounds and variables. Compartments or parts of the dialogue may be played back or used in an actual or simulated environment until the application is proven and ready for use. Thereafter, the dialogue flow can be converted into a machine readable code used by interactive voice response systems with a minimum requirement for manual or machine software implementation.

2. Sharp

Sharp discloses a device for controlling a sound reproduction transcriber which provides alternate playback intervals and stopped intervals. The playback interval length in words and the average word reproduction rate are manually selectable by a secretary or instructor. The device meters the reproduced words during the playback interval and initiates the stopped interval at the first audio pause after the selected number of words have been played back. The required total cycle time is computed by effectively dividing the number of words played back during the playback interval by the selected average word reproduction rate. A timer, which counts the time elapsed since the playback interval was initiated, has its counted time compared to the

computed total cycle time so that the stopped interval is terminated, the circuit is reset and the playback interval is again initiated upon coincidence of the computed time and the elapsed time.

B. Traversal of the Rejections on the Art

Applicant's originally filed independent claims require the loading of a user simulation script the specifies voice interactions with a "voice application". The independent claims further require the derivation from that voice application of a nominal output to include either a text stream or an audio stream. The independent claims yet further require the processing of the user simulation script to generate a simulated output for the voice application which corresponds to the nominal output. Finally, the independent claims require the calculation of an execution time for the simulated output, the execution time being equal to a length of the text stream divided by an empirical speaking rate of a user and the execution of the simulated output at its calculated execution time in conjunction with the voice application.

Exemplary claim 1 recites as follows:

> A method for simulating a run-time user interaction with a voice application, said method comprising the steps of:
>
> loading a user simulation script programmed to specify simulated voice interactions with the voice application;
>
> deriving from the voice application a nominal output, the nominal output including a text stream;
>
> processing the user simulation script to generate a simulated output for the voice application corresponding to the nominal output;
>
> calculating an execution time for the simulated output, the execution time being equal to a length of the text stream divided by an empirical speaking rate of a user; and
>
> executing the simulated output at its calculated execution time in conjunction with the voice application.

Thus, as originally filed, claim 1 (and through similar language, claims 5 and 9) require the derivation of a nominal output from a voice application where the nominal output is an audio stream in claims 3, 7 and 11).

Figure 3 of Hank illustrates the refinement of a speech corpus using text provided by a "speech assistant". As defined in column 4, lines x - x, the "speech assistant" is a tool used to produce a vendor specific speech grammar. The excerpt describing "speech assistant in Hanks follows verbatim:

> Further, the same information in text form in the DA may be presented via line 51 to the Speech Assistant (SA) 52 which the designer uses via keyboard 41 to produce vendor specific ASR grammar at block 53 via line 54 or to produce vendor specific ASR grammar in run time format in block 55 via line 56 or to employ the grammar compiler 57 to change the information in block 53 to a vendor specific ASR grammar in block 55.

The "speech corpus" shown in Figure 3, in turn, is undefined throughout Hanks, though it is generally accepted that a speech corpus is a database of speech audio files and text transcriptions in a format that can be used to create acoustic models which can then be used with a speech recognition engine.

In all of the Applicants' claims, however, a "nominal output" must be derived from a "voice application" which is shown in paragraph [0020] of the Applicants' specification to mean, "[A]ny logic permitting user interaction through a voice driven user interface, such as a mark-up language specification for voice interaction with some form of coupled computing logic." Importantly, as defined in paragraph [0028] of Applicants' specification, a "nominal output" is an output that the voice application would otherwise provide to a user in a non-simulated environment. Thus, it is not possible for the nominal output to be the speech corpus itself as suggested by the Examiner. As a result, considering both the widely accepted definition of "speech corpus" in conjunction with the actual teaching of Figure 3 and the Applicants' adoption

of the common industry meaning for "voice application", the Applicants believe that the referenced portion of Hank, namely Figure 3 and column 4, lines 23 through 34 do not teach "deriving from a voice application a nominal output including either a text stream (claims 1, 5, 9) or an audio stream (claims 3, 7, 11) because Figure 3 does not show the derivation of a text stream or the derivation of an audio stream from a "voice application".

Of note, Sharp fails to cure the deficiencies of Hank as Sharp does not relate to the same field of endeavor as Hank (Hank is about grammar generating tools and Sharp is about a transcription device). Further, Sharp definitely does not disclose the limitation "calculating an execution time for the simulated output, the execution time being equal to a length of the text stream divided by an empirical speaking rate of a user" because the Examiner has inappropriately equated the "empirical speaking rate of a user" with a "word reproduction rate" that is selected by a switch on an apparatus. Of course, it is impossible for Sharp to teach "executing the simulated output at its calculated execution time in conjunction with the voice application" because Sharp wholly lacks a "voice application" in which to execute simulated output.

IV. Conclusion

M.P.E.P. 2143 entitled the Basic Requirements of a Prima Facie Case of Obviousness state with particularity,

> To establish a prima facie case of obviousness, three basic criteria must be met. First, there must be some suggestion or motivation, either in the references themselves or in the knowledge generally available to one of ordinary skill in the art, to modify the reference or to combine reference teachings. Second, there must be a reasonable expectation of success. Finally, **the prior art reference (or references when combined) must teach or suggest all the claim limitations.**

In that several of the limitations of all of the claims are lacking in the combination of Hank and Sharp, the Applicants respectfully request the withdrawal of the rejections under 35 U.S.C. § 103(a) owing to the clearly distinctive nature of a "voice application" and "nominal output" as compared to a "speech assistant" and "speech corpus" and the foregoing remarks. The Applicants request that the Examiner call the undersigned if clarification is needed on any matter within this Amendment, or if the Examiner believes a telephone interview would expedite the prosecution of the subject application to completion.

Respectfully submitted,

Date: August 17, 2007

/Steven M. Greenberg/

Steven M. Greenberg, Reg. No.: 44,725
Attorney for Applicant(s)
Carey, Rodriguez, Greenberg & Paul, LLP
950 Peninsula Corporate Circle, Suite 3020
Boca Raton, Florida 33487
Customer No. 46322
Tel: (561) 922-3845
Fax: (561) 244-1062